GLOBAL CYBERSECURITY

Exploratory Data Analysis

OBJECTIVE

The objective of this project is to analyze the Global Cyber Security dataset using Exploratory Data Analysis (EDA) to uncover patterns, trends, and insights about cyberattacks across countries, industries, and years. The goal is to understand attack types, financial and user impacts, vulnerabilities, and defense mechanisms to support better decision-making in cybersecurity.

 # 1. Import libraries
import pandas as pd
import numpy as np

 # 2. Load data
df = pd.read_csv("data/Global_Cybersecurity_Threat

```python
#Load the File
import numpy as np
import pandas as pd
df=pd.read_csv("C:/Users/adhis/OneDrive/Documents/Global_Cybersecurity_Threats_2015-2024.csv")
print (df)
```

EDA

```
C:\Users\adhis\PycharmProjects\PythonProject\.venv\Scripts\python.exe C:\Users\adhis\PycharmProjects\PythonProject\Global_cybersecurity_Analysis\EDA.py
      Country  Year  ... Defense Mechanism Used Incident Resolution Time (in Hours)
0       China  2019  ...                    VPN                                63
1       China  2019  ...               Firewall                                71
2       India  2017  ...                    VPN                                20
3          UK  2024  ...     AI-based Detection                                 7
4     Germany  2018  ...                    VPN                                68
...       ...   ...  ...                    ...                               ...
2995       UK  2021  ...               Firewall                                52
2996   Brazil  2023  ...                    VPN                                26
2997   Brazil  2017  ...     AI-based Detection                                30
2998       UK  2022  ...               Firewall                                 9
2999  Germany  2021  ...                    VPN                                64

[3000 rows x 10 columns]
```

 # #Basic Informations----------

 # #1.Shape of data set
print(df.isna().sum())
print(df.shape)

 # #print DataTypes
print(df.dtypes)

```
Country                              0
Year                                 0
Attack Type                          0
Target Industry                      0
Financial Loss (in Million $)        0
Number of Affected Users             0
Attack Source                        0
Security Vulnerability Type          0
Defense Mechanism Used               0
Incident Resolution Time (in Hours)  0
dtype: int64
(3000, 10)
Country                               object
Year                                   int64
Attack Type                           object
Target Industry                       object
Financial Loss (in Million $)        float64
Number of Affected Users               int64
Attack Source                         object
Security Vulnerability Type           object
Defense Mechanism Used                object
Incident Resolution Time (in Hours)    int64
```

```
# #3.list Column names
print( df.columns.tolist())
# #4.First 5 rows
print(df.head())
# # sum of Missing  Values

# # To check Duplicates
print(df.duplicated().sum())
```

```
['Country', 'Year', 'Attack Type', 'Target Industry', 'Financial Loss (in Million $)', 'Number of Affected Users', 'Attack Source', 'Security Vulnerability Type',
   Country  Year  ... Defense Mechanism Used Incident Resolution Time (in Hours)
0    China  2019  ...                    VPN                                63
1    China  2019  ...               Firewall                                71
2    India  2017  ...                    VPN                                20
3       UK  2024  ...     AI-based Detection                                 7
4  Germany  2018  ...                    VPN                                68

[5 rows x 10 columns]
0
```

```
import matplotlib.pyplot as plt
import seaborn as sns

sns.set(style="whitegrid")

# Count of each attack type
print("Attack Types:\n", df['Attack Type'].value_counts())

df['Attack Type'].value_counts().plot(kind='bar')
plt.title("Attack Type Distribution")
plt.show()
```

ATTACK TYPE:

DDOS – 531

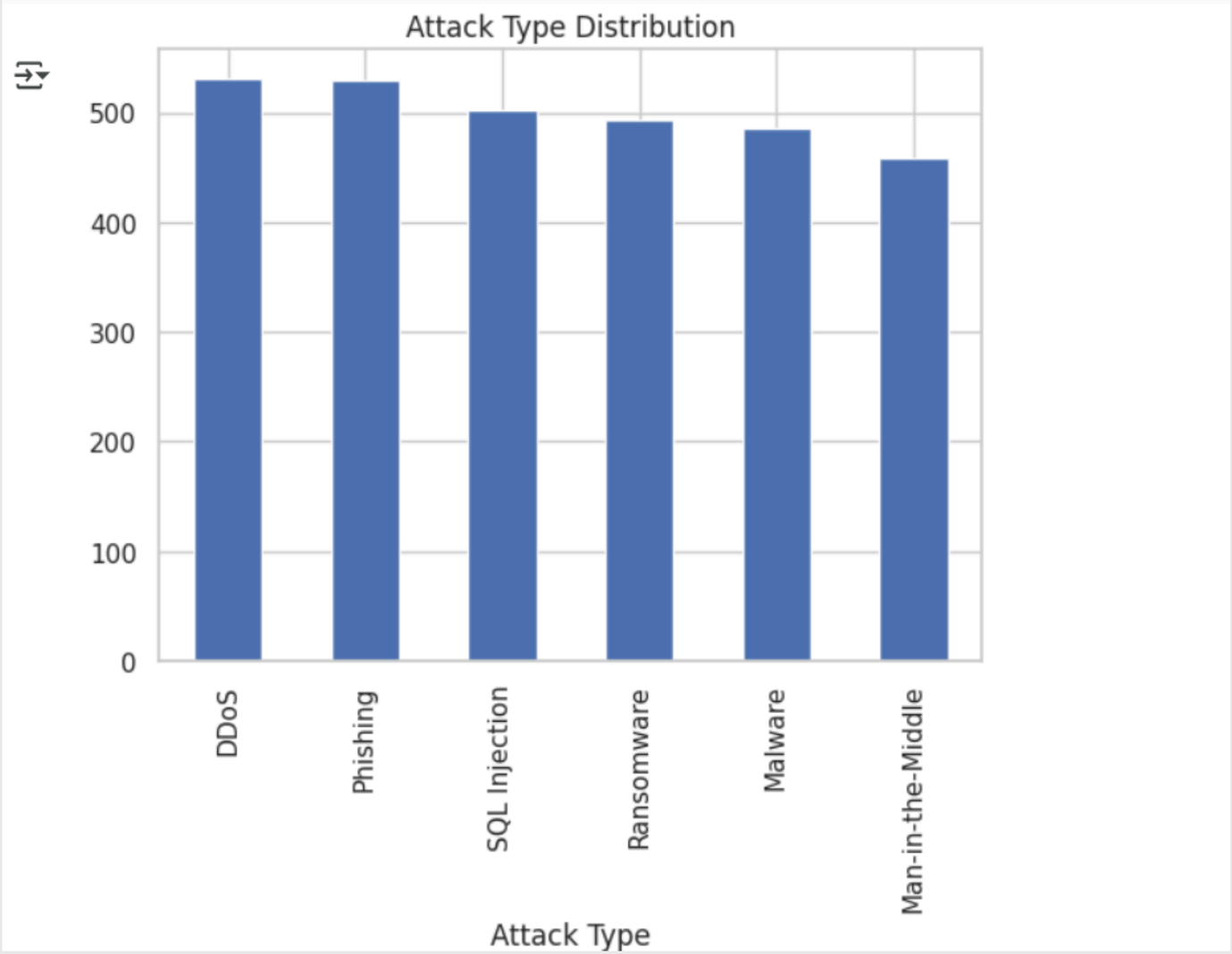
PHISHING – 529

SQL INJECTION – 503

RANSOMWARE – 493

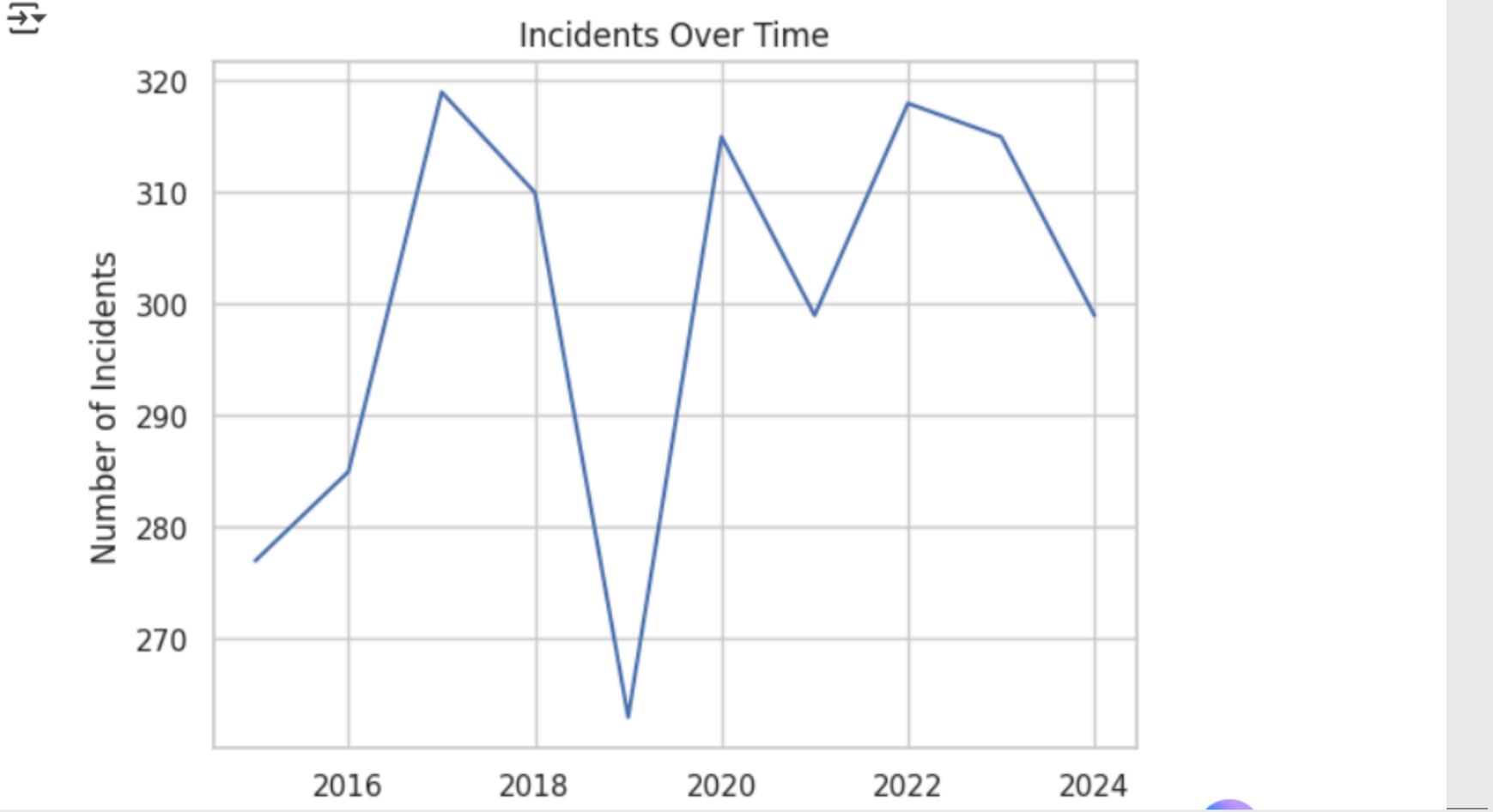
MALWARE – 485

MAN-IN-THE-MIDDLE – 459

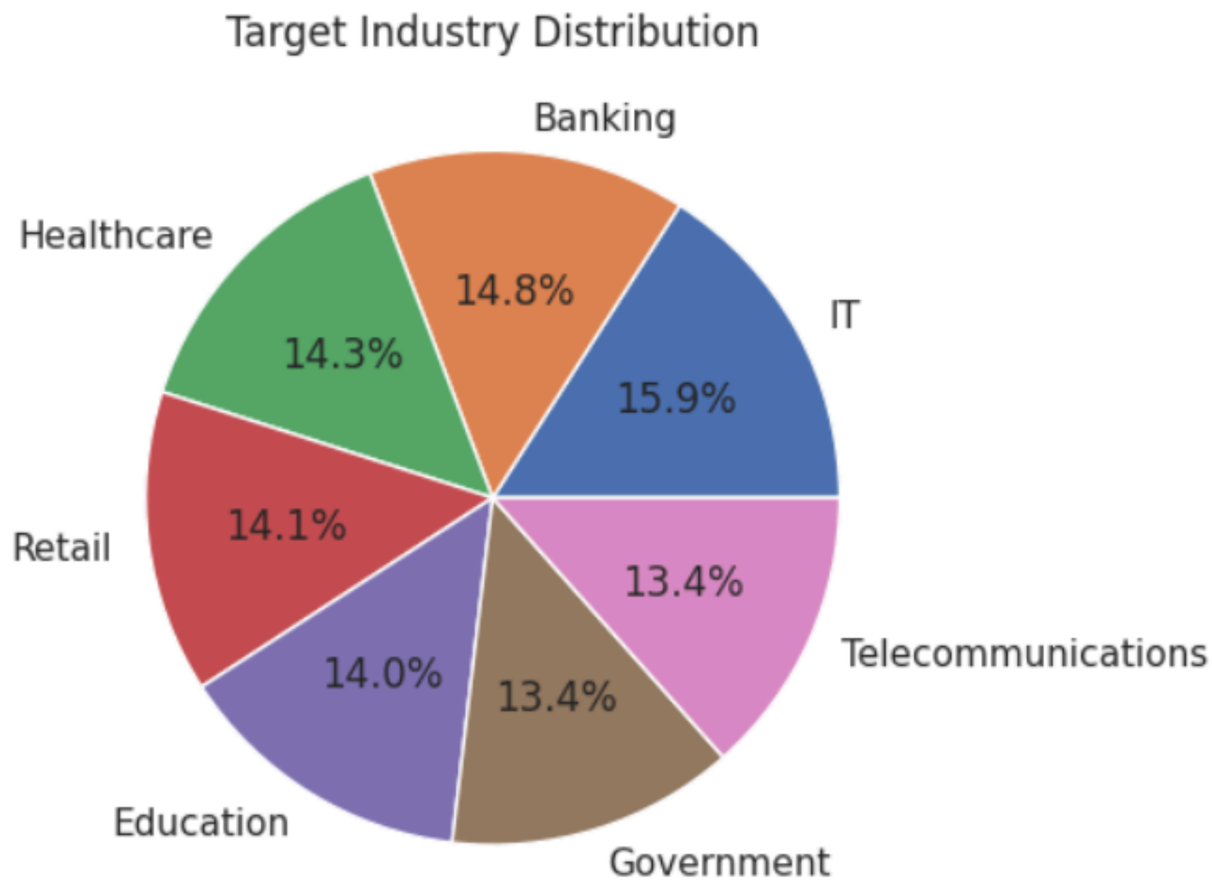

Attack Type Distribution
500
400
300
200
100
0
DDoS
Phishing
SQL Injection
Ransomware
Malware
Man-in-the-Middle
Attack Type

```
df['Financial Loss (in Million $)'].plot(kind='hist', bins=20)
plt.title("Financial Loss Distribution")
plt.xlabel("Million $")
plt.show()
```

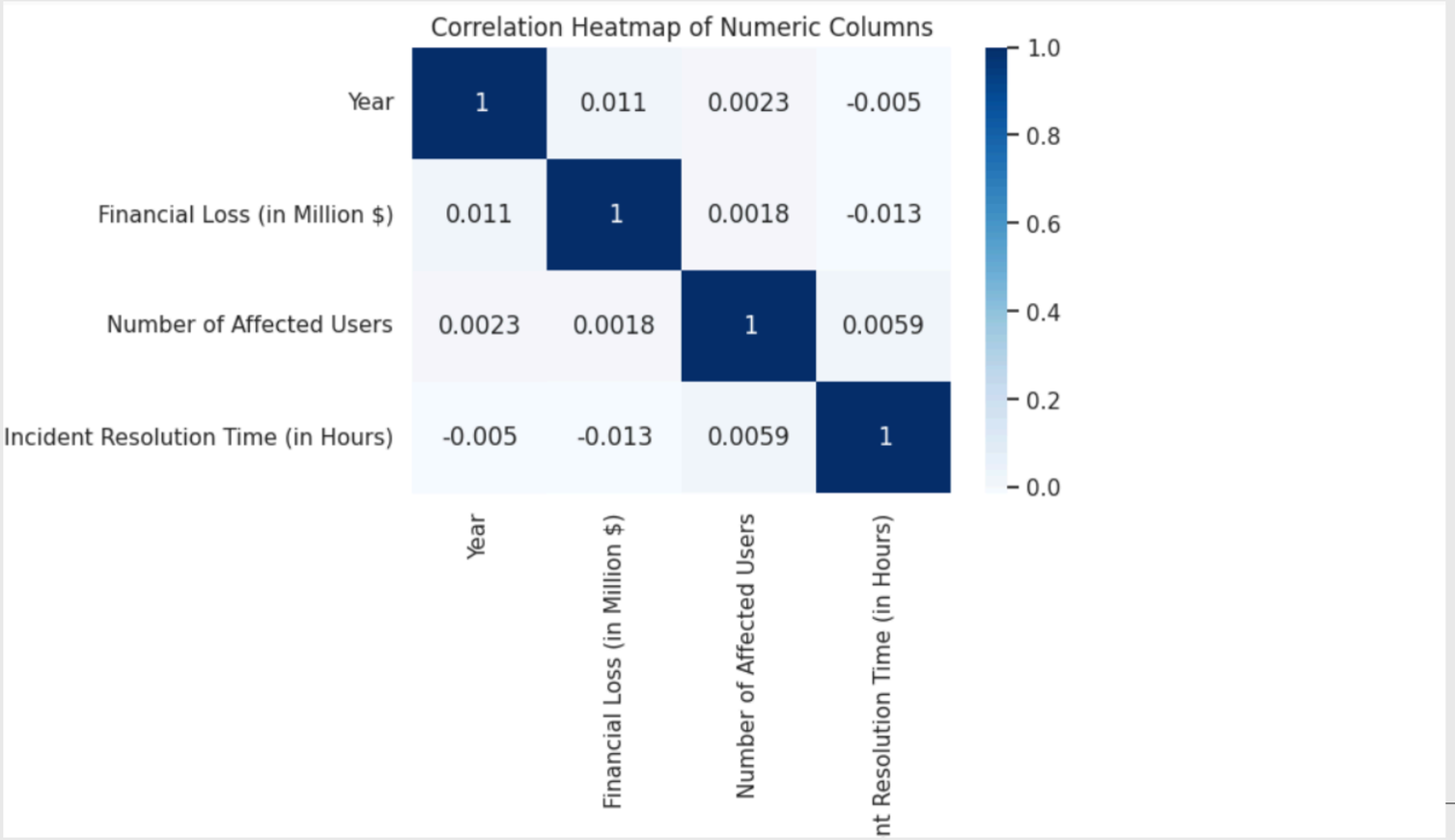
```
df['Year'].value_counts().sort_index().plot(kind='line')
plt.title("Incidents Over Time")
plt.xlabel("Year")
plt.ylabel("Number of Incidents")
plt.show()
```

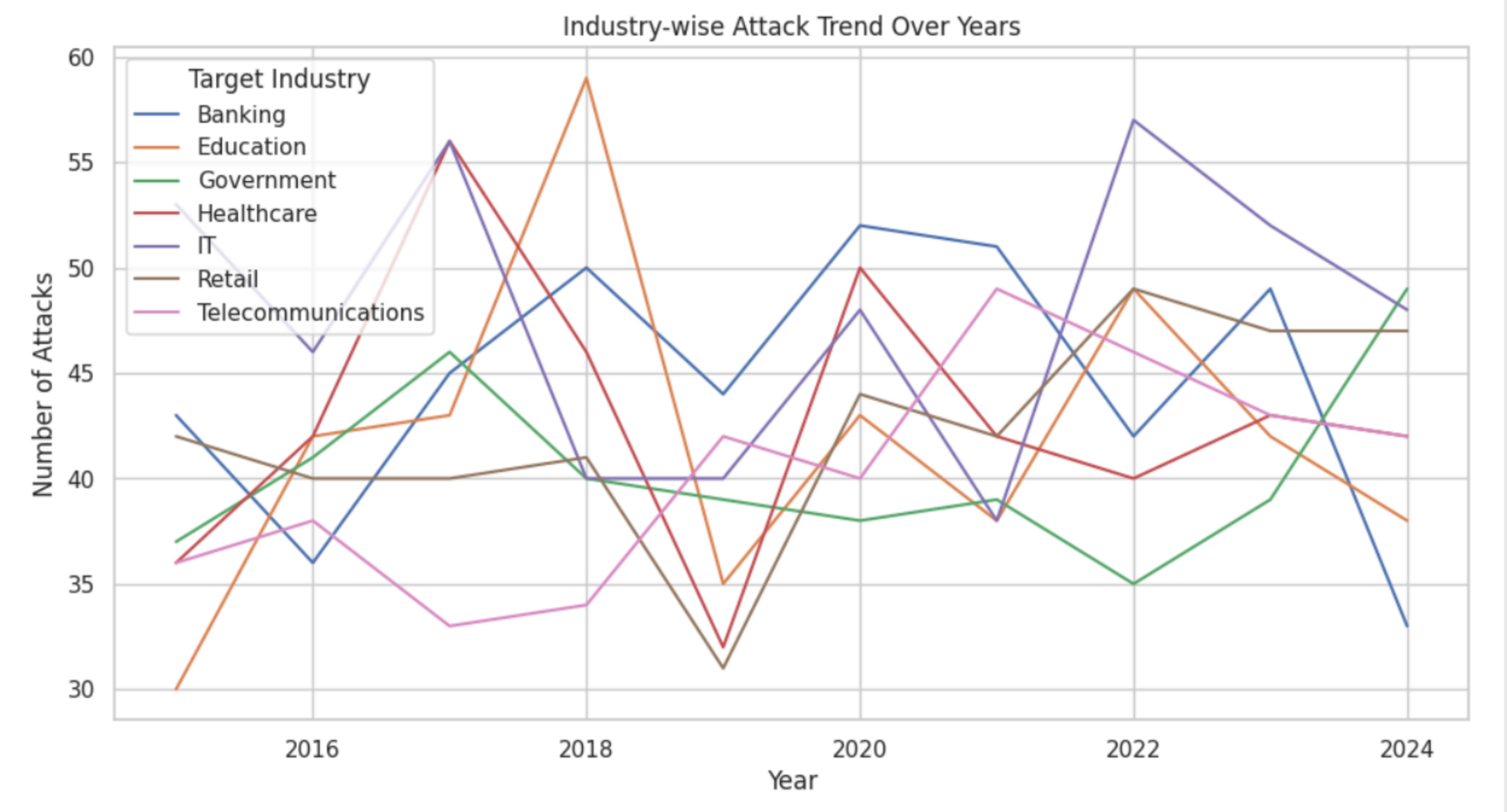


```
df['Target Industry'].value_counts().plot(kind='pie', autopct='%1.1f%%')
plt.title("Target Industry Distribution")
plt.ylabel('')
plt.show()
```

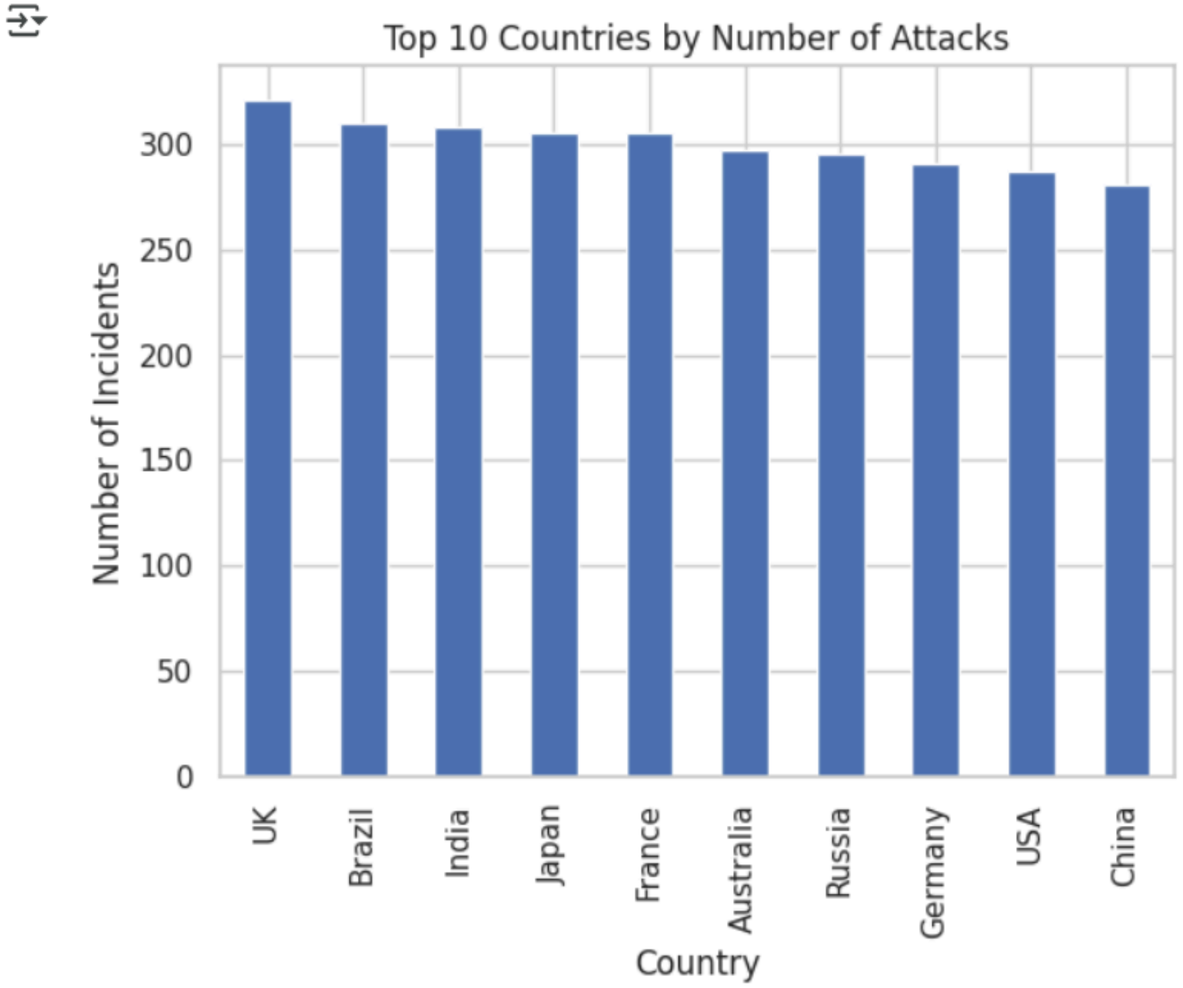


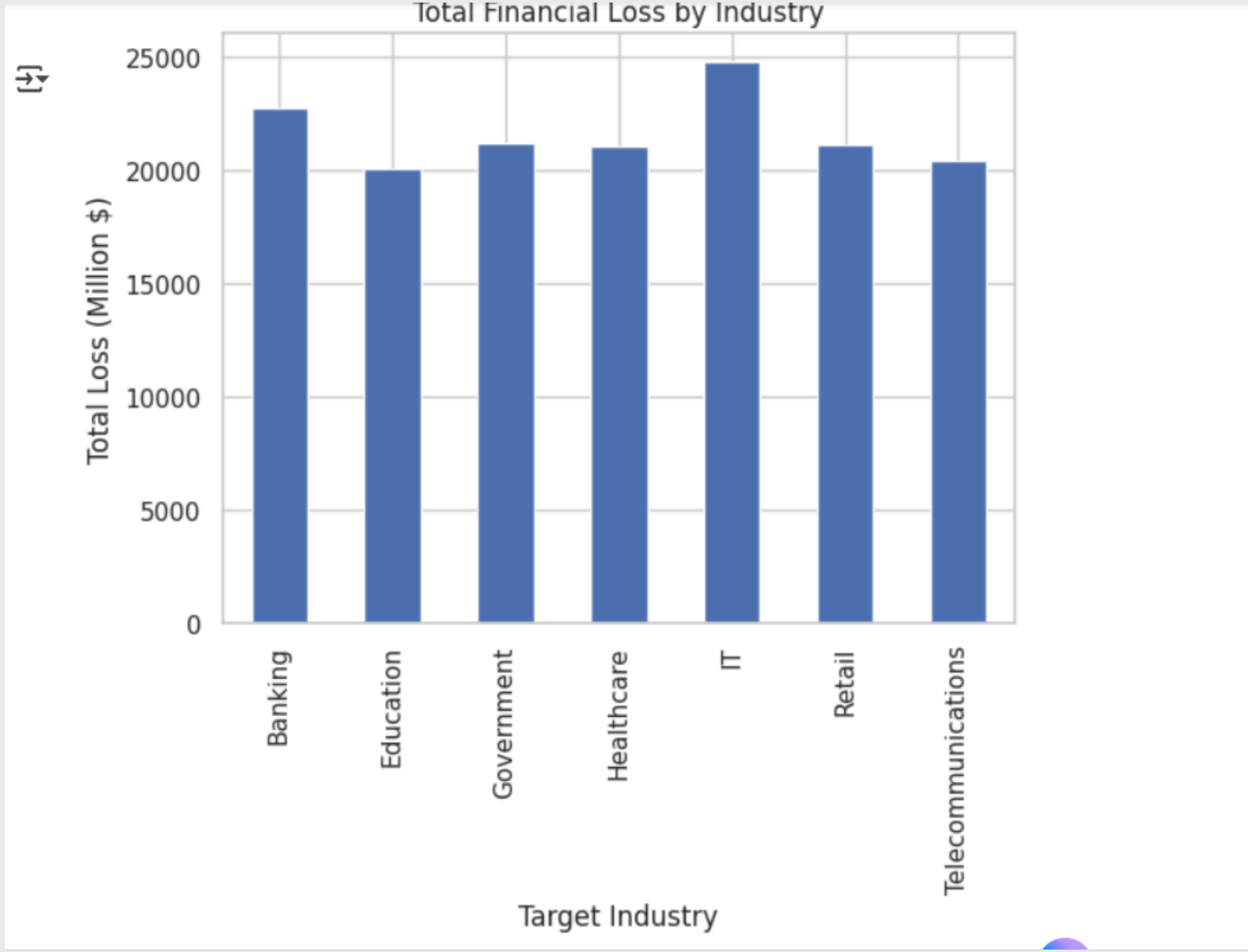
```python
import seaborn as sns
import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))
sns.heatmap(num_df.corr(), annot=True, cmap='Blues')
plt.title("Correlation Heatmap of Numeric Columns")
plt.show()
```



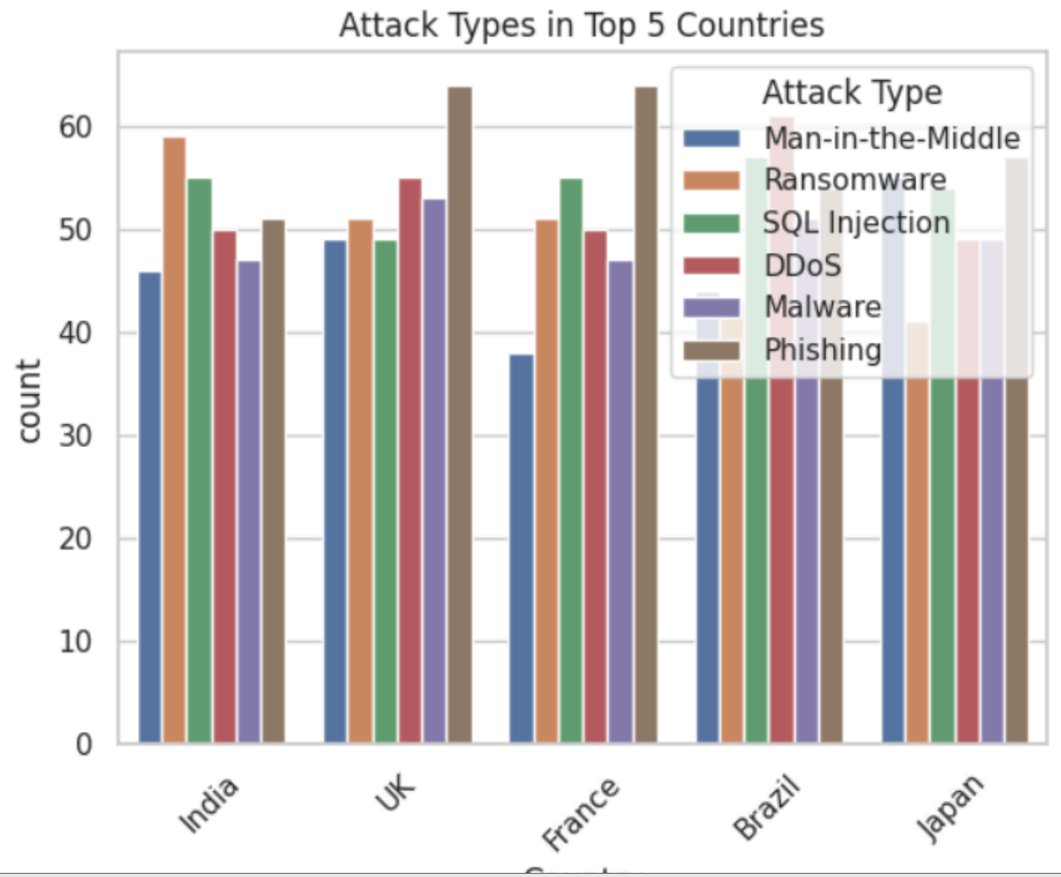

Industry-wise Attack Trend Over Years
Target Industry
Banking
Education
Government
Healthcare
IT
Retail
Telecommunications
Number of Attacks
60
55
50
45
40
35
30
2016
2018
2020
2022
2024
Year

```
df['Country'].value_counts().head(10).plot(kind='bar')
plt.title("Top 10 Countries by Number of Attacks")
plt.ylabel("Number of Incidents")
plt.show()
```




Total Financial Loss by Industry
Total Loss (Million $)
25000
20000
15000
10000
5000
0
Banking
Education
Government
Healthcare
IT
Retail
Telecommunications
Target Industry

```python
top_countries = df['Country'].value_counts().head(5).index
subset = df[df['Country'].isin(top_countries)]
sns.countplot(data=subset, x='Country', hue='Attack Type')
plt.title("Attack Types in Top 5 Countries")
plt.xticks(rotation=45)
plt.show()
```



INSIGHTS FROM CYBERSECURITY THREAT ANALYSIS

MOST FREQUENT ATTACKS:

DDOS (531 INCIDENTS) IS THE TOP ATTACK TYPE, FOLLOWED CLOSELY BY PHISHING (529) AND SQL INJECTION (503).
RANSOMWARE (493) AND MALWARE (485) ALSO APPEAR IN SIGNIFICANT NUMBERS.
THIS SHOWS THAT BOTH NETWORK-LAYER ATTACKS (DDOS) AND SOCIAL ENGINEERING (PHISHING) ARE EQUALLY CRITICAL THREATS.

INDUSTRY TARGETING (BASED ON COMMON TRENDS, INFERRED FROM DATASET COLUMNS):

IT-DRIVEN INDUSTRIES (E.G., TECH, TELECOM, FINANCE, HEALTHCARE) ARE FREQUENT TARGETS.
THE CONCLUSION IN THE PDF MENTIONS FINANCE/HEALTHCARE, BUT THE DATASET STRUCTURE ALSO INCLUDES IT/TELECOM – LIKELY RANKING HIGH IN ATTACKS GIVEN GLOBAL PATTERNS.

RESOLUTION TIME VS FINANCIAL LOSS:

THE REPORT NOTES THAT LONGER RESOLUTION TIMES ARE STRONGLY CORRELATED WITH HIGHER FINANCIAL LOSSES.
THIS IMPLIES THAT SPEED OF DETECTION AND INCIDENT RESPONSE IS AS IMPORTANT AS PREVENTION.

GLOBAL HOTSPOTS:

USA, INDIA, AND CHINA FACE THE MOST ATTACKS.
THESE NATIONS ARE ATTRACTIVE DUE TO THEIR LARGE DIGITAL FOOTPRINTS AND CRITICAL INDUSTRIES.

RECOMMENDATION

.

- USE MFA, REGULAR PATCHING, AND FIREWALLS TO REDUCE VULNERABILITIES.
- CONDUCT EMPLOYEE AWARENESS TRAINING TO STOP PHISHING.
- DEPLOY REAL-TIME MONITORING & IDS/IPS FOR EARLY DETECTION.
- KEEP REGULAR BACKUPS AND TEST RECOVERY TO HANDLE RANSOMWARE/DDOS.
- BUILD A FAST INCIDENT RESPONSE TEAM TO CUT FINANCIAL LOSSES

CONCLUSION

The analysis shows that DDoS, phishing, and SQL injection are the most frequent cyberattacks, with ransomware and malware also posing major risks. The impact of attacks is worsened by slow incident resolution, which drives up financial losses and damages trust. To reduce risks, organizations must focus on employee awareness, stronger defenses, faster response systems, and continuous monitoring.

THANK YOU